

## Getting Started with PSS Data

### Slide 1 of 23

Title Slide: Getting Started with PSS Data

### Slide 2 of 23

This module introduces users to the data collected through the 2011-12 PSS. It describes the process of obtaining previously tabulated data and generating custom tables from EISi and the resources that are available to learn more about the surveys, the data, and the data files.

The module describes the contents of the 2011-12 PSS data file, its variables, and variable naming conventions. Information presented in this module will be helpful for understanding some of the more detailed information presented in subsequent PSS modules. For this reason, users who are planning to proceed through the modules and use PSS data for analytic purposes are strongly encouraged to complete this module first.

### Slide 3 of 23

Information in PSS was collected from each private school identified in the study frame. The questionnaire is designed for completion by a single respondent who could have been the principal, assistant or vice principal, or any other administrative employee. PSS data were collected by online questionnaires, by paper questionnaires or by telephone interviews. The data collection instrument is available online and can be accessed by clicking the 'Data Files' link on the PSS homepage or the underlined screen text, 'paper questionnaire.'

The next module in this series – titled 'Data Collected Through the Private School Universe Survey' – describes the range of topics covered in PSS in more detail and provides information about the frame and derived variables that are available for analysis. The module can be accessed by clicking the corresponding underlined screen text.

### Slide 4 of 23

Public-use descriptive statistics are survey results that have been generated by NCES. You must use a statistical software package to weight the data for proper analysis. Public-use data files allow researchers to manipulate the data, but have been altered to prevent inadvertent disclosure of respondent identity. The most easily accessed PSS data results - that is, public-use descriptive statistics - are only available in PSS reports, the PSS Table Library, or by generating tables within the Elementary / Secondary Information System, or EISi. Each of these sources of PSS data results is accessible by clicking the corresponding underlined screen text.

The public-use data are available in a variety of formats for download from the NCES website. EISi or published reports may not cover all of the variables of interest and the public-use data file does not contain the number of high-school graduates from the

## Getting Started with PSS Data

previous year. More information can be obtained from the restricted-use files, which will be discussed later in this module.

### Slide 5 of 23

A report is available for the 2011-12 PSS. This report can be accessed via the corresponding underlined screen text. It can also be accessed from the PSS homepage of the NCES website by clicking the 'Publications & Products' link, then the 'Reports' link, and then selecting the appropriate report. This report features a few data elements from PSS, but only covers a fraction of the topics available from the collected PSS data. The first part of the report presents a page of "selected findings" or highlights of data results, as well as a number of data tables with more detailed results. Following the data tables and standard error tables, there is a technical summary of PSS data collection, data processing, response rates, and imputation and weighting methods.

### Slide 6 of 23

The PSS Table Library is a set of tables with data from 2011-12 PSS as well as some previous rounds of PSS. These tables are organized by survey year. To see all of the tables generated for a given year, users can click on the year of interest. More data and analysis options are available using EISi, which is available by selecting the 'Data Tools' link on the PSS homepage.

### Slide 7 of 23

Users can also generate customized data results or tables from the public-use PSS data by using EISi. The Elementary/Secondary Information System (or EISi) is an NCES web application that allows users to view public and private school data, and create custom tables and charts from PSS data. To access the system, click 'Data Tools' and then 'EISi.'

### Slide 8 of 23

Once accessed, EISi can be used to search published tables and create custom tables and analyses. On the screen, examples of the options available within EISi including *quickFacts*, *expressTables*, and *tableGenerator* are provided. Let's look at the *quickFacts* link first.

### Slide 9 of 23

After accessing 'EISi *quickFacts*,' select 'Private School' as the level. After selecting the level, select a characteristic, then select a year and a school to generate custom data. For example, after selecting '*quickFacts*,' select 'private school' as the level and "total enrollment" as the characteristic. Now the user can type in the name of a school and choose from the drop down box. '*quickFacts*' is useful to quickly view single data elements for a particular private school.

## Getting Started with PSS Data

### Slide 10 of 23

With EISi 'expressTables' you may view and sort most requested data tables at the private school level. After selecting 'Private School' as the level, select a table for the column variables and then filters appropriate to your specific analyses. For example, select "private school characteristics" as the table and select the state and county before selecting "view table."

### Slide 11 of 23

To create custom tables for private schools, use the 'EISi *tableGenerator*.' For example, select "private school" as the table row and "2011-2012" as the years. Next, select the desired table columns and additional filters before selecting "create table."

### Slide 12 of 23

The PSS public-use data files include variables providing information obtained from the questionnaire. Data files for multiple years are available in SAS, SPSS, and TXT format. Each public-use data file is accompanied by documentation that includes the file layout, the questionnaire, and the codebook. Public-use data sets require the use of a statistical software package for proper analysis of weighted data.

The public-use data files can be accessed by clicking the corresponding underlined screen text.

### Slide 13 of 23

Restricted-use datasets allow researchers to analyze survey results with full access to the data supplied by respondents. As discussed in the DLDT Common Module titled, "Acquiring micro-level NCES datasets," restricted-use data are only available to researchers who apply for and are granted a restricted-use license to use them. The restricted-use files for a given round of data collection contain all of the original survey responses and variables. Researchers who are uncertain of which data file is right for their analytic purposes should first examine the public-use descriptive statistics. Restricted-use data sets require the use of a statistical software package for proper analysis of weighted data.

### Slide 14 of 23

The decision whether to seek a restricted-use license agreement with NCES depends on whether the basic data tables provided satisfy your research needs. Exploring EISi for the data PSS data of interest may allow researchers quick access to the data needed for analysis. It's important to keep in mind, though, that while many data tables can be generated using EISi, it does not allow for variable construction or row variable selection. With the public-use or restricted-use data files, users can manipulate response categories and combine items into new variables.

## Getting Started with PSS Data

### Slide 15 of 23

If restricted-use PSS data are needed to address the research question of interest, the Restricted-Use Data Procedures Manual should be downloaded from the NCES website. The manual can be accessed by clicking the corresponding underlined screen text. The manual provides step-by-step instructions on the process of applying online for the restricted-use license. It is not possible to download restricted-use data files from any part of the NCES website. To access the Restricted-use Licensing web page, please click on the corresponding underlined screen text.

### Slide 16 of 23

Several important resources are available within the restricted-use DVD for researchers who are using restricted-use data. Access the README file first, as it is the best place to begin learning about these data and the procedures for accessing the data. The README file contains information about the DVD contents and structure. It also provides you with installation information. The data files are in ASCII, SAS, and SPSS format on the DVD. ASCII allows “raw” data to be read into any type of statistical software program without imposing any formatting. This allows importation of the raw data into any statistical software program. Because Excel does not allow the data to be weighted, it is not an appropriate choice for research with PSS data.

The DVD also contains PDFs of the PSS questionnaires. The questionnaires provide the source code variable name for each questionnaire item, along with the exact item wording and the response categories. Much more complete detail on how the survey instruments were designed, how the survey was conducted, how the data were processed, and how the data files can be merged, is contained in the Data File User's Manual.

### Slide 17 of 23

The questionnaire variables in the PSS data file are named using a standard convention. While the naming conventions for other variables such as imputation flags, weights, and frame or derived variables will be discussed later in this module, for questionnaire variables, the first character of the questionnaire variable name is the letter “P.” For questionnaire variables only, the first character of the questionnaire variable name is the letter “P.” The rest of the variable name indicates the numeric source code that is found next to the questionnaire item on the survey instrument. Source codes are three digits and are not necessarily in numeric order on the questionnaire but are put in numeric order on the data files. Each source code is assigned to a particular questionnaire item response. There may be one or more source codes for any given questionnaire item. This is why downloading the PDF of the questionnaire is so important – you need to see the exact survey item wording, what the item response categories are and obtain source codes for the variables of interest to you. For example, item 6 has a source code of 305. On the data file, the variable name for this item is P305.

## Getting Started with PSS Data

### Slide 18 of 23

Variable labels offer more description than the variable names. Key words from the item are incorporated in an abbreviated form into the variable label. For example, let's consider the variable 'P305' that is associated with item 6 from the questionnaire. The 'P' in the first character position indicates that this variable is a questionnaire variable, and the number is the source code for item 6. The variable label associated with 'P305' is 'Total Student Enrollment.'

### Slide 19 of 23

Variables were classified as frame variables if they were drawn from the 2009–11 PSS file or based on the 2010–11 PSS frame update. Examples of frame variables include the school's permanent identification number (or PPIN) or the ZIP code of the school's mailing address (or PZIP). Derived variables are re-coded variables, or variables based on more than one questionnaire item. These variable names are entirely alphabetic, with no source code numbers. They were created for users using information from two or more variables, two or more sources, or both. Derived variables have been created for reporting purposes, and to help analysts save time and effort by not having to recreate commonly-used variables themselves. Detailed descriptions of all the derived variables can be found in the data file documentation.

### Slide 20 of 23

PSS data are fully imputed – this means all data records contain a response for each questionnaire item. As a result, each item from the PSS questionnaires will be represented by two variables in the PSS data files: the fully imputed questionnaire variable and the imputation flag variable. Flags are used to indicate, or flag, which imputations were conducted for each variable individually. Imputation flag variables all start with "F underscore" followed by the variable name. In other words, "F\_P250" is the imputation flag corresponding to the questionnaire item 250. When an item was skipped, or no response was provided, the imputation flag value was set to 4 or 5, depending on the method used to impute the value.

The imputation flag is set to zero when a valid response was provided to the actual questionnaire item, meaning no imputations were necessary. The imputation flag is set to 4 if the item was imputed using hot deck imputation. Hot deck imputation is selecting a value from a matched respondent. Hot deck methods impute missing values within a data matrix by using available values from the same matrix. The object, from which these available values are taken for imputation within another, is called the donor. The imputation flag is set to 5 if the item was imputed based upon post-imputation review.

### Slide 21 of 23

The selection of the geographic areas in the dual frame approach described within the module titled, 'Introduction to the Private School Universe Survey' introduces a probabilistic element to the PSS, which is accounted for by sampling weights for the area frame cases. It is important to note that even though PSS is a universe study,

## Getting Started with PSS Data

analysts will need to use sampling weights to account for the probabilistic nature of the area frame and the nonresponse adjustment.

The Final Weight (PFNLWGT) incorporates the probabilities of selection, as well as the nonresponse adjustment needed to bring the sample estimate into the correct proportion for the target population. The replicate weights (REPW1 through REPW88) are a set of weights that are needed to calculate standard errors, which will be described in more detail in the module titled “PSS Weights, Variance, and Missing Data.”

### Slide 22 of 23

In addition to the resources already discussed, there are two additional resources that may be of interest. The NCES Online Bibliography Search Tool allows you to see what other researchers have published using PSS data. To use the Search Tool, at a minimum, you must enter in the Data Source (PSS is one of the NCES datasets listed). You do not have to know the exact research article title or the author’s name, as any field left blank widens the search results. The Data File User’s Manual for the particular survey year has more in-depth coverage of topics such as frame creation, data collection and response rates, data preparation, guide to the data files and codebooks, and user notes and cautions. Each of these resources can be accessed by clicking the corresponding underlined screen text.

### Slide 23 of 23

This module has introduced you to the data collected across the 2011-12 PSS. It described the process of obtaining previously tabulated data and generating custom tables from EISi and the resources that are available to learn more about the survey, the data, and the data file.

Additionally, important resources that have been provided throughout the module are summarized here along with the module’s objectives for your reference.

You may now proceed to the next module in the series or exit the module.